Gender Recognition from Face Images Using a Fusion of SVM Classifiers

George Azzopardi[1], Antonio Greco[2(✉)], and Mario Vento[2]

[1] University of Malta, Msida, Malta
george.azzopardi@um.edu.mt
[2] University of Salerno, Fisciano, Italy
{agreco,mvento}@unisa.it

Abstract. The recognition of gender from face images is an important application, especially in the fields of security, marketing and intelligent user interfaces. We propose an approach to gender recognition from faces by fusing the decisions of SVM classifiers. Each classifier is trained with different types of features, namely HOG (shape), LBP (texture) and raw pixel values. For the latter features we use an SVM with a linear kernel and for the two former ones we use SVMs with histogram intersection kernels. We come to a decision by fusing the three classifiers with a majority vote. We demonstrate the effectiveness of our approach on a new dataset that we extract from FERET. We achieve an accuracy of 92.6 %, which outperforms the commercial products Face++ and Luxand.

Keywords: Gender recognition · HOG · LBP · Histogram intersection

1 Introduction

In the last few years the identification of certain demographic attributes, such as gender, age and race, has involved various research areas, including computer vision. Here we concentrate on the automatic gender recognition problem from face images. Such a system has various applications. Examples include behaviour adaptation of intelligent user interfaces and adaptive advertising billboards. A system that stores the gender of each customer could help to collect demographic statistics in order to evaluate the effectiveness of the marketing strategies. Furthermore, video surveillance systems could first use gender identification in order to reduce the search space in their database, yielding a much more efficient retrieval process.

The gender recognition task seems to be effortless for humans. In fact, an individual is able to distinguish a male from a female by simply observing the face [1]. It also seems that the visual system of our brains has developed neurons that are selective for faces [2]. The detection of the facial features, such as beard or mustache, eyes, earrings, make-up among others, could be very challenging for a computer vision algorithm. These problems are mainly due to the variations in pose, facial expressions, occlusions, and changes in illumination. Other challenges

A. Campilho and F. Karray (Eds.): ICIAR 2016, LNCS 9730, pp. 533–538, 2016.
DOI: 10.1007/978-3-319-41501-7_59

include the intra-variation within the two classes due to age and racial features. Nevertheless, most of the methods [3–5] perform gender recognition by using only features that are extracted from faces.

A typical facial gender recognition algorithm carries out four steps: (i) face detection; (ii) pre-processing; (iii) feature extraction; and (iv) binary classification. Most of the approaches use the Viola-Jones algorithm [6] for face detection and the Support Vector Machine [7] for classification. The main difference lies in the type of features extracted from faces. The most used features for gender recognition are the pixel intensity values, the Local Binary Pattern (LBP) descriptor and the fiducial distances. The latter are the distances between specific facial points (e.g. eye corners, face contour, tip of the nose) and have been widely used in recent years for face recognition [5,8].

The pixel intensity values, also called raw features, can be directly used to train a binary classifier for gender recognition. Sometimes, the raw feature extraction is carried out after a pre-processing step, in order to deal with pose (face alignment) and brightness variations (histogram equalization). Techniques such as principal component analysis (PCA) have also been used to reduce the dimensionality of the feature vectors [3,9,10].

The histograms of LBP features [11] in the face region are often used as discriminative feature vectors for gender classification [12]. Other methods that combine LBP and shape features were proposed in [4,13], where pixel intensity values were also considered.

A different approach is based on the search of the facial landmarks, the so-called fiducial points, and their mutual spatial arrangement. These points could be labeled by hand or using predefined masks like the Active Shape Model (ASM) [14]. The distances between these fiducial points are used as discriminant features for gender recognition [8]. Deep learning and convolutional neural networks have also been proposed for the localization of facial landmarks [5].

The contribution of our work is two-fold. First, we take the majority vote from the output of three SVM classifiers that are configured with different types of features, namely HOG [15], LBP and raw pixels. We demonstrate that shape, texture and raw features are orthogonal to each other as we achieve considerable higher accuracy when combining them. Second, we evaluate our method on a subset of the FERET [16], a standard benchmark dataset for face recognition, which we collected specifically for gender recognition purposes. We made it publicly available with the name GENDER-FERET [17], thus our results are reproducible and can be compared with new approaches.

2 Method

Figure 1 shows the architecture of the proposed method. First we perform face detection using the Viola-Jones algorithm [6]. Then we crop the detected face and resize it to 128×128 pixels. We transform the image into a $(128 \times 128 =)$ 16384-element feature vector and divide each element by 255 so that all dimensions have the same range of [0,1].

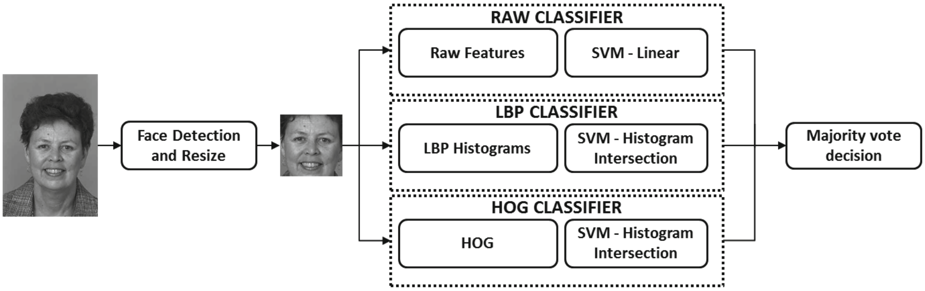

Fig. 1. Architecture of the proposed method

We apply the LBP descriptor [11] to the entire image by comparing the intensity value of each pixel with a 3×3 neighbourhood. We use a spatial tiling of 3×3 and generate a 256-element L2-normalized histogram for each tile. Finally, we merge the nine histograms to form a $(256 \times 9 =)$ 2304-element vector for each image. We use the LBP histogram-based descriptors with an SVM classifier characterized by a histogram intersection kernel.

As to the HOG descriptor we first compute the gradient and angle of every pixel by considering the responses of first-order partial derivatives of a 2D Gaussian function with a $\sigma = 1$. Then we sample blocks of 32×32 pixels that overlap by 50 % and for each block we use a spatial tiling of 2×2. For each tile we compute the L2-normalized weighted histogram of 9 bins (in intervals of 20°), clip the normalized values at 0.2 and normalize again. Considering we use face images of size 128×128 pixels, the HOG descriptor results in a (7 blocks $\times$ 7 blocks $\times$ 4 tiles $\times$ 9 bins $=$) 1764-element vector. Similar to the LBP-based descriptor, since this descriptor is based on histograms, we train an SVM classifier with a histogram intersection kernel.

The result of each classifier is a pair of probabilities that a given image has a male or a female face. In order to come up with a decision we sum the three male probabilities and the three female probabilities and if the total male probability is greater than the total female probability then we label the given face image to be a male otherwise a female.

3 Evaluation

Despite the effort of the Benchmarking Facial Image Analysis Technologies (BeFIT) [18], there is not yet a standard dataset for the evaluation of gender recognition algorithms. Most of the available datasets are designed for face detection and recognition. FERET [16] is among the most important benchmark datasets for face recognition, but it does not provide gender annotations. For this reason, we extracted a subset of FERET by choosing only the frontal images, without variations in pose, but with different expressions, backgrounds and illumination conditions, as shown with some examples in Fig. 2. The dataset is balanced, so we

Fig. 2. Examples of faces in the GENDER-FERET dataset. The bounding boxes indicate the faces detected by Viola-Jones [6]. The images in the last column are (top) the average male face and (bottom) the average female face of this dataset.

Table 1. Experimental results. The first 5 columns show the combination of features that we use. Every set of features is combined with an SVM classifier whose type of kernel, Linear or Histogram Intersection (H.Int) is specified underneath. The last five columns show the results in terms of male (M) and female (F) true positives (TP) and false positives (FP), along with accuracy.

Raw (Linear)	LBP (Linear)	HOG (Linear)	LBP (H.Int)	HOG (H.Int)	M (TP)	F (TP)	M (FP)	F (FP)	Accuracy (%)
✓					208	209	28	27	88.3
	✓				205	201	31	35	86.0
		✓			204	209	32	27	87.5
✓	✓				215	212	21	24	90.5
✓		✓			212	214	24	22	90.3
	✓	✓			212	211	24	25	89.6
✓	✓	✓			215	214	21	22	90.9
			✓		189	213	47	23	85.2
				✓	208	217	28	19	90.0
✓			✓		212	214	24	22	90.3
✓				✓	219	215	17	21	91.9
			✓	✓	212	220	24	16	91.5
✓			✓	✓	217	220	19	16	**92.6**
Face++ [19]					233	190	3	46	89.6
Luxand [20]					235	186	1	50	89.2

have the same number of male and female images (473 m, 473 f). Then we selected 50 % of the images to form a training set (237 m, 237 f) and the remaining 50 % to form a test set (236 m, 236 f). The face of a person is either in the training or in the test set, but not in both. Different variations in expression, facial features, background and illumination were also considered in both sets. We call the newly formed dataset GENDER-FERET and make it publicly available [17]. We use the accuracy rate as a performance measurement on the test set.

As shown in Table 1, the accuracy is above 85 % using any combination of feature or classifier. The classifier that relies on only raw pixel values achieves 88.3 % of accuracy, which is better than both the LBP- and the HOG-based classifiers. Performance significantly increases when we combine the decisions of different classifiers. Using only linear SVMs the highest accuracy that we achieve is 90.9 %, while the best accuracy (92.6 %) is achieved using histogram intersection kernel for both the LBP and HOG classifiers.

Since there are no standard datasets for the gender recognition algorithms, the comparison with the results published in other studies is infeasible. For a more realistic comparison on the same test set, we apply the commercial libraries, namely Face++ [19] and Luxand [20]. In Table 1 we also report the results of these commercial methods, which are substantially lower than the performance achieved by our method.

4 Discussion and Conclusion

We propose a method for gender recognition from face images using a fusion of SVM classifiers. The method has been evaluated on the new GENDER-FERET dataset, which includes variations in expressions, facial features, background and illumination. The experimental results show that by taking the majority vote of three SVM classifiers we are able to significantly improve the performance of gender recognition from face images. The result suggests that the HOG-, LBP-, and pixel-based descriptors, which essentially describe the properties of shape, texture and intensity distribution, are complementary features. By using the majority vote rule, we can exploit the abilities of the three classifiers, and come to a more certain decision.

The confusion matrices reported in Table 1 show that our method has a good generalization ability, while Face++ and Luxand seem to be specialized in the recognition of males. It is important to note that we could not train the Face++ and Luxand algorithms on the same training set that we used to configure our system. This is because only the preconfigured versions of these algorithms are available. In future work we will use larger and heterogeneous datasets in order to evaluate the generalization ability of the classifier and to compare the performance of the proposed method in even more challenging situations.

References

1. Marquardt Beauty Analysis. Face variations by sex (2014). http://www.beautyanalysis.com/beauty-and-you/face-variations/face-variations-sex/
2. Perrett, D.I., Rolls, E.T., Caan, W.: Visual neurones responsive to faces in the monkey temporal cortex. Exp. Brain Res. **47**(3), 329–342 (1982)
3. Moghaddam, B., Yang, M.-H.: Learning gender with support faces. IEEE Trans. Pattern Anal. Mach. Intell. **24**(5), 707–711 (2002)

4. Alexandre, L.A.: Gender recognition: a multiscale decision fusion approach. Pattern Recognit. Lett. **31**(11), 1422–1427 (2010)
5. Sun, Y., Wang, X., Tang, X.: Deep convolutional network cascade for facial point detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3476–3483 (2013)
6. Viola, P., Jones, M.J.: Robust real-time face detection. Int. J. Comput. Vis. **57**(2), 137–154 (2004)
7. Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. **20**(3), 273–297 (1995)
8. Brunelli, R., Poggio, T.: Face recognition: features versus templates. IEEE Trans. Pattern Anal. Mach. Intell. **10**, 1042–1052 (1993)
9. Baluja, S., Rowley, H.A.: Boosting sex identification performance. Int. J. Comput. Vis. **71**(1), 111–119 (2007)
10. Yang, J., Zhang, D., Frangi, A.F., Yang, J.-Y.: Two-dimensional PCA: a new approach to appearance-based face representation and recognition. IEEE Trans. Pattern Anal. Mach. Intell. **26**(1), 131–137 (2004)
11. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)
12. Lian, H.-C., Lu, B.-L.: Multi-view gender classification using local binary patterns and support vector machines. In: Wang, J., Yi, Z., Żurada, J.M., Lu, B.-L., Yin, H. (eds.) ISNN 2006. LNCS, vol. 3972, pp. 202–209. Springer, Heidelberg (2006)
13. Tapia, J.E., Perez, C.A.: Gender classification based on fusion of different spatial scale features selected by mutual information from histogram of LBP, intensity, shape. IEEE Trans. Inf. Forensics Secur. **8**(3), 488–499 (2013)
14. Milborrow, S., Nicolls, F.: Locating facial features with an extended active shape model. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part IV. LNCS, vol. 5305, pp. 504–513. Springer, Heidelberg (2008)
15. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005, vol. 1, pp. 886–893. IEEE (2005)
16. Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J.: The FERET evaluation methodology for face-recognition algorithms. IEEE Trans. Pattern Anal. Mach. Intell. **22**(10), 1090–1104 (2000)
17. Mivia Lab University of Salerno. Gender-FERET dataset (2016). http://mivia.unisa.it/database/gender-feret.zip
18. Karlsruhe Insitute of Technology. Befit - benchmarking facial image analysis technologies (2011). http://fipa.cs.kit.edu/412.php
19. Face++. Leading face recognition on cloud (2014). http://www.faceplusplus.com/
20. Luxand. Facial feature detection technologies (2015). https://www.luxand.com/